

ABSTRACT

A system, method, and computer program for automatically generating text analysis systems is disclosed. Individual passes of a multi-pass text analyzer are created by generating rules from samples supplied by users. Successive passes are created in a cascading fashion by performing partial text analyses employing existing passes. A complete text analyzer interleaves the generated passes with a framework of existing passes. The complete text analysis system can then process texts to identify patterns similar to samples added by users. Generation of rules from samples encompasses a wide range of constructs and granularities that occur in text, from individual words to intrasentential patterns, to sentential, paragraph, section, and other formats that occur in text documents.

004290-96840960